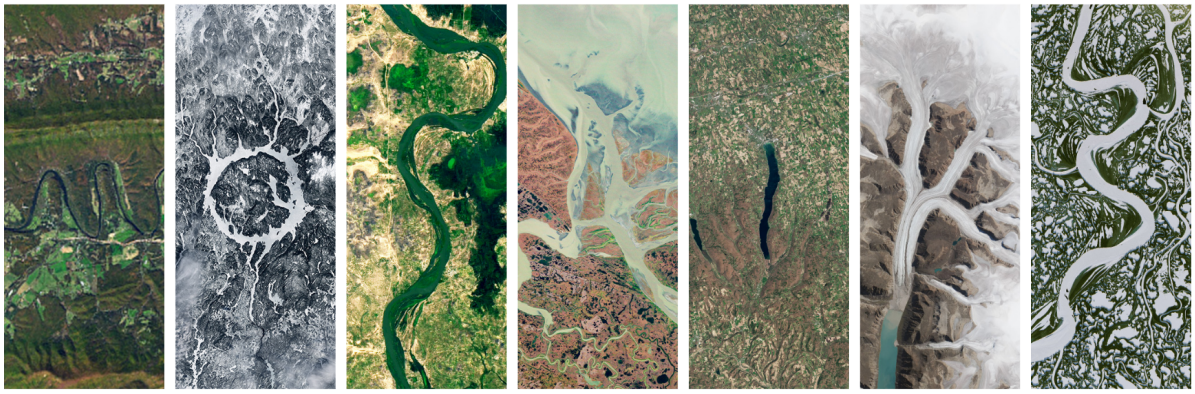
MOSAIKS Training Manual

Cullen Molitor Tamma Carleton Esther Rolf

Table of contents

Welcome

This is the first ever MOSAIKS Training Course! This book will serve as a reference for learning the ins and outs of MOSAIKS throughout the training.



Made with: Your Name in Landsat

MOSAIKS Stands for **Multi-task Observation using SAtellite Imagery & Kitchen Sinks**. It is a framework that aims to simplify using satellite imagery and machine learning to answer questions about socioeconomic and environmental outcomes across different geographic contexts and time periods.

This comprehensive two-week program is designed for academics, professionals, and policy makers interested in leveraging MOSAIKS for socioeconomic and environmental outcomes.

This course covers the fundamentals of working with MOSAIKS, from basic concepts to advanced applications. The training is particularly suited for those working in:

- Remote sensing and satellite imagery analysis
- Machine learning applications with geospatial data
- Agricultural monitoring and assessment
- Development research and policy making

Throughout this course, you'll learn about satellite imagery processing, MOSAIKS feature extraction, uncertainty quantification, and practical applications of MOSAIKS in real-world scenarios. We'll combine theoretical knowledge with hands-on

exercises, ensuring you gain both conceptual understanding and practical experience.

The curriculum includes working with various data sources, processing satellite imagery, understanding Random Convolutional Features (RCFs), implementing machine learning models, interpreting results, and applying predictive models in various contexts.

You will also explore important considerations in using MOSAIKS for survey data, particularly relevant for development research applications.

Whether you're new to MOSAIKS or looking to deepen your expertise, this course will provide you with the tools and knowledge needed to effectively utilize this powerful framework in your work.

Part I

Introduction

Course structure

This course is designed as an intensive two-week program that combines lectures, demonstrations, and hands-on sessions. Each day is structured as follows:

Time	Activity
9:00 - 10:30	Morning Session 1
10:30 - 11:00	Break
11:00 - 12:30	Morning Session 2
12:30 - 1:30	Lunch
1:30 - 3:00	Afternoon Session 1
3:00 - 3:30	Break
3:30 - 4:30	Afternoon Session 2
4:30 - 5:00	Feedback and Development

Each day concludes with a Q&A and feedback session from 4:30-5:00, providing opportunities to clarify concepts and share ideas. It is expected that this first course will spur many new ideas and concepts which should be included in the following trainings. Please remember to take notes throughout each day with particular emphasis in areas you think could be explained better for future classes. These can be areas that you struggled with or that you would anticipate could be difficult for others.

Schedule overview

Week 1

- **Monday**: Orientation and introduction to MOSAIKS
- **Tuesday**: Ground labels and data processing fundamentals
- **Wednesday**: Agriculture applications and MOSAIKS API
- **Thursday**: Satellite imagery fundamentals and processing
- **Friday**: Deep dive into Random Convolutional Features (RCFs)

Week 2

- **Monday**: Martin Luther King Jr. Day (holiday - no class)
- **Tuesday**: Task modeling and machine learning applications
- **Wednesday**: Understanding uncertainty in MOSAIKS

- **Thursday**: Survey data processing and design
- **Friday**: Future directions and advanced applications

Training expectations

What you will learn

- Understanding of MOSAIKS framework and capabilities
- Practical skills in satellite imagery processing
- Experience with machine learning applications
- Hands-on practice with real-world datasets
- Knowledge of survey data integration
- Best practices for model implementation

Prerequisites

There are no explcicit prerequisites, though this course does cover some advanced topics in:

- The `Python` programming language
- Machine Learning
- Geospatial data

Participant expectations

- Active participation in discussions and hands-on sessions
- Completion of assigned homework (particularly the Week 1 Friday assignment)
- Engagement in Q&A sessions
- Contribution to feedback sessions for course improvement

Computing requirements

The course includes hands-on computing sessions. You will need:

- A computer with access to the internet
- A Google account
- Acess to Google Colaboratory
- Access to necessary data (details to be provided)

Homework and presentations

There will be a homework assignment at the end of Week 1, which participants will present on Tuesday of Week 2. This assignment is designed to reinforce learning and provide practical experience with MOSAIKS tools.

1 Compute Setup

This course will primarily use Google Colaboratory (Colab) for our computational needs. Colab is a free, cloud-based platform that allows you to write and execute Python code through your browser. It comes with many pre-installed libraries and provides free access to computing resources, including GPUs.

1.1 Requirements

To participate in the coding portions of this course, you'll need:

- A laptop or desktop computer
- Reliable internet connection
- A Google account (if you don't have one, create one at accounts.google.com)
- A modern web browser (Chrome recommended)

1.2 Getting Started with Google Colab

1.2.1 Accessing Colab

1. Go to colab.research.google.com
2. Sign in with your Google account
3. Click "New Notebook" to create your first Colab notebook

1.2.2 Understanding the Interface

The Colab interface is similar to Jupyter notebooks, with a few key components:

- **Menu Bar**: Contains File, Edit, View, Insert, Runtime, Tools, and Help options
- **Toolbar**: Quick access to common actions like adding code/text cells
- **Cell Area**: Where you write and execute code or text
- **Runtime Status**: Shows the state of your notebook's connection to Google's servers

1.2.3 Basic Operations

1. **Creating Cells**:
    - Code cells: Click "+ Code" or use Ctrl+M B
    - Text cells: Click "+ Text" or use Ctrl+M M
2. **Running Cells**:
    - Click the play button next to the cell
    - Use Shift+Enter
    - Select Runtime > Run all from the menu
3. **Cell Types**:
    - Code cells: For Python code execution
    - Text cells: For documentation (supports Markdown)

1.2.4 Important Features

1. **Runtime Type**:
    - Click Runtime > Change runtime type
    - Select Python 3 as the runtime
    - For GPU access: Change hardware accelerator to GPU when needed
2. **File Management**:
    - Files uploaded to Colab are temporary
    - Connect to Google Drive for persistent storage:

```
from google.colab import drive
drive.mount('/content/drive')
```

1. **Package Installation:**

Install additional packages using:

1.3 conda

```
# add a -c conda-forge to select the conda-forge channel
# add a -q flag to install quietly (reduced output)
# add a -y flag to prememptively accept other changes
!conda install package_name
```

1.4 pip

```
!pip install package_name
```

1.4.1 Best Practices

1. Save Your Work:
    - Regularly save to Google Drive (File > Save a copy in Drive)
    - Download important notebooks locally as backups
2. Resource Management:
    - Close unused notebooks to free up resources
    - Be aware of idle timeouts (notebooks disconnect after extended inactivity)
3. Memory Usage:
    - Monitor memory usage through Runtime > Resource usage
    - Use Runtime > Factory reset runtime if you run into memory issues

1.4.2 Keyboard Shortcuts

Here are the most useful keyboard shortcuts for working in Colab:

1.5 Windows/Linux

Action	Shortcut
Run current cell	Ctrl+Enter
Run cell and move to next	Shift+Enter
Run cell and insert below	Alt+Enter
Insert code cell above	Ctrl+M A
Insert code cell below	Ctrl+M B
Convert to text cell	Ctrl+M M
Convert to code cell	Ctrl+M Y
Delete current cell	Ctrl+M D
Toggle line numbers	Ctrl+M L
Toggle output	Ctrl+M O
Cut cell	Ctrl+M X
Copy cell	Ctrl+M C
Paste cell below	Ctrl+M V
Select multiple cells	Shift+Up/Down
Find and replace	Ctrl+F
Save notebook	Ctrl+S

1.6 Mac

Action	Shortcut
Run current cell	⌘+Enter
Run cell and move to next	Shift+Enter
Run cell and insert below	Option+Enter
Insert code cell above	⌘+M A
Insert code cell below	⌘+M B
Convert to text cell	⌘+M M
Convert to code cell	⌘+M Y
Delete current cell	⌘+M D
Toggle line numbers	⌘+M L
Toggle output	⌘+M O
Cut cell	⌘+M X
Copy cell	⌘+M C
Paste cell below	⌘+M V
Select multiple cells	Shift+Up/Down
Find and replace	⌘+F
Save notebook	⌘+S

You can view all available shortcuts in Colab by pressing Ctrl+M H (⌘+M H on Mac) or through Help > Keyboard shortcuts in the menu.

1.6.1 Common Issues and Solutions

1. Runtime Disconnections:
    - Click "Reconnect" when prompted
    - Your variables will be reset, but saved code remains
2. Package Installation Issues:
    - Restart runtime after installing new packages
    - Use Runtime > Restart runtime
3. Memory Errors:
    - Clear unnecessary variables
    - Restart runtime
    - Consider using smaller data samples during development

1.6.2 Getting Help

- Access Colab's built-in documentation: `Help > Colab Overview`
- View keyboard shortcuts: `Help > Keyboard shortcuts`
- Search the Help menu for specific topics
- Use the `Help > Search Solutions` feature

[Note: This would be a good place to add screenshots showing key interface elements and operations]

In the next sections of this course, we'll be using Colab extensively for hands-on exercises. Make sure you're comfortable with these basics before proceeding.

1.7 Accessing Course Notebooks

All course notebooks are hosted on GitHub and can be accessed directly in Google Colab. There are two ways to access the notebooks: